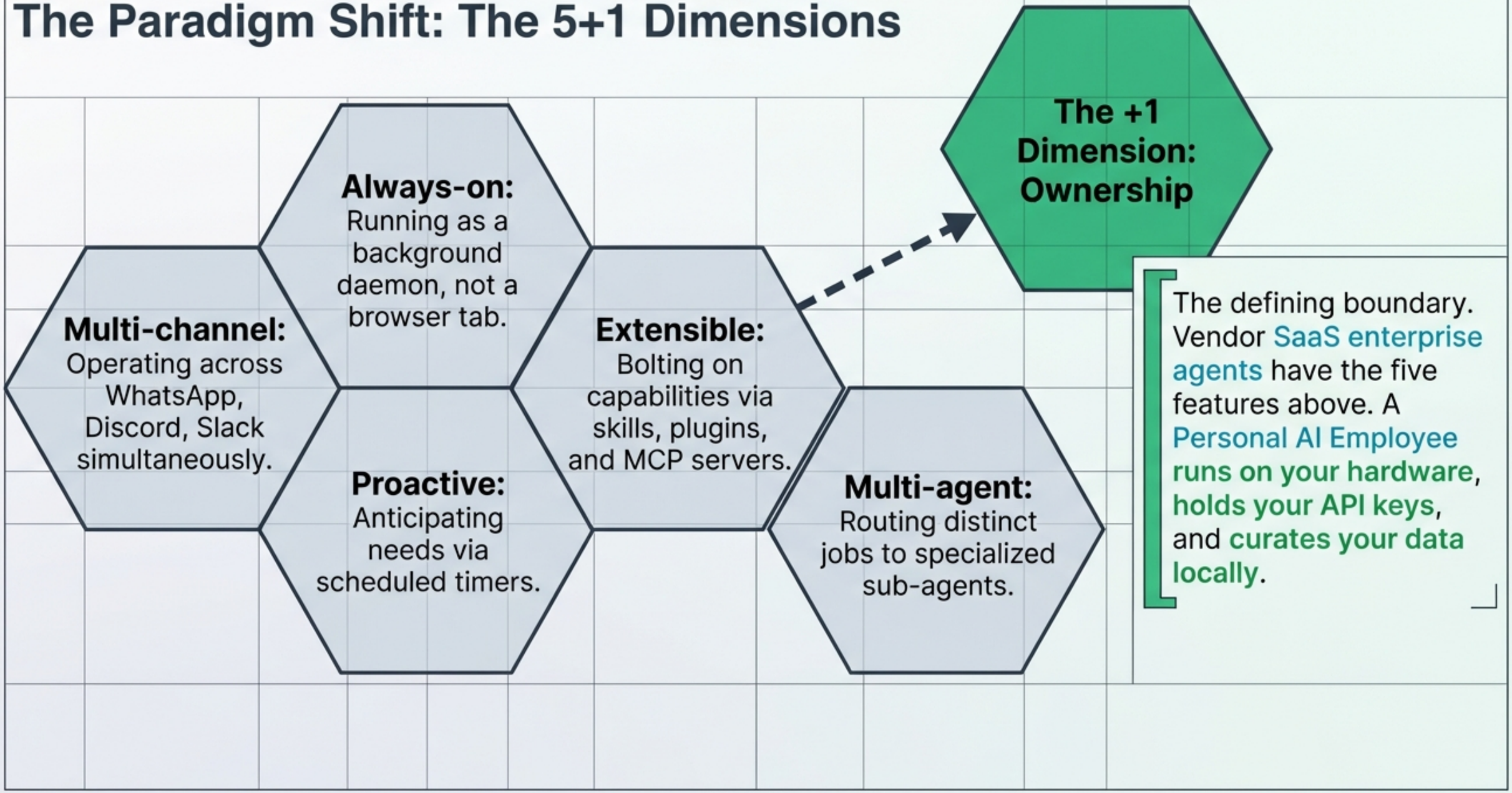
OpenClaw Technical Blueprint

Architecting the Personal AI Employee

An engineering synthesis of deployment realities, state management, and zero-trust security architectures.

The Paradigm Shift: The 5+1 Dimensions

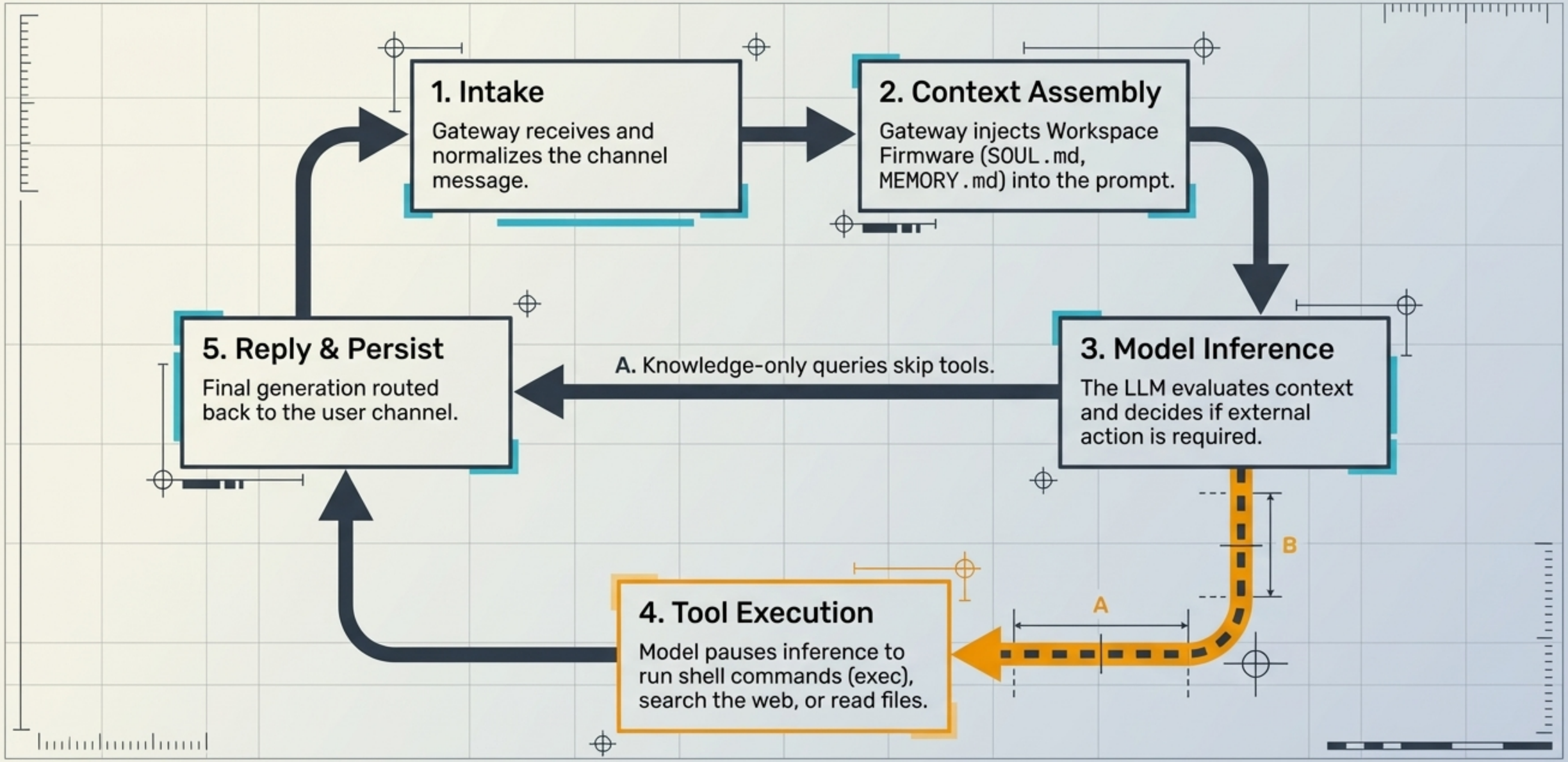


The defining boundary. Vendor SaaS enterprise agents have the five features above. A Personal AI Employee **runs on your hardware**, **holds your API keys**, and **curates your data locally**.

The Agent OS Mental Model

OpenClaw Component	OS Analogue	Functional Definition
Gateway	Kernel	Routes messages, manages sessions, coordinates plugins.
Workspace Files	Firmware	Reloads per-message; defines personality, identity, and behavioral boundaries (SOUL.md).
Plugins	Device Drivers	Adds native capabilities (channels, voice, tooling).
Sessions	Process Memory	Holds per-conversation context, strictly isolated per user.
Heartbeats / Crons	Cron Daemon	Background periodic awareness and precise execution.
Channels	IPC / Networking	I/O adapters for external messaging platforms.

The Execution Engine: 5-Step Agent Loop

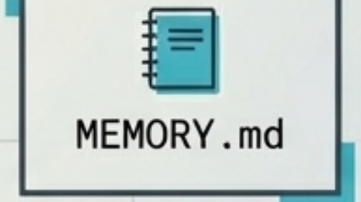


State & Memory Architecture

Volatile: Session Memory

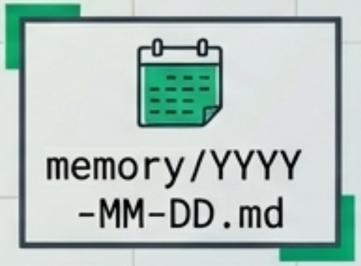
Per-conversation context. Fades completely when the session ends or resets.

Persistent: Workspace Memory



`MEMORY.md`

The **curated notebook**. Loads entirely at the start of every session. Best for core user preferences and permanent facts.



`memory/YYYY-MM-DD.md`

The **daily journal**. Auto-loads today and yesterday. Older logs are retrieved dynamically via hybrid vector/keyword **memory_search**.

Instant State Checks: Slash Commands >_

`/commands bypass the model entirely.`
`Intercepted by the Gateway for zero-token, instant state checks.`
`Examples: /status, /model, /reset`

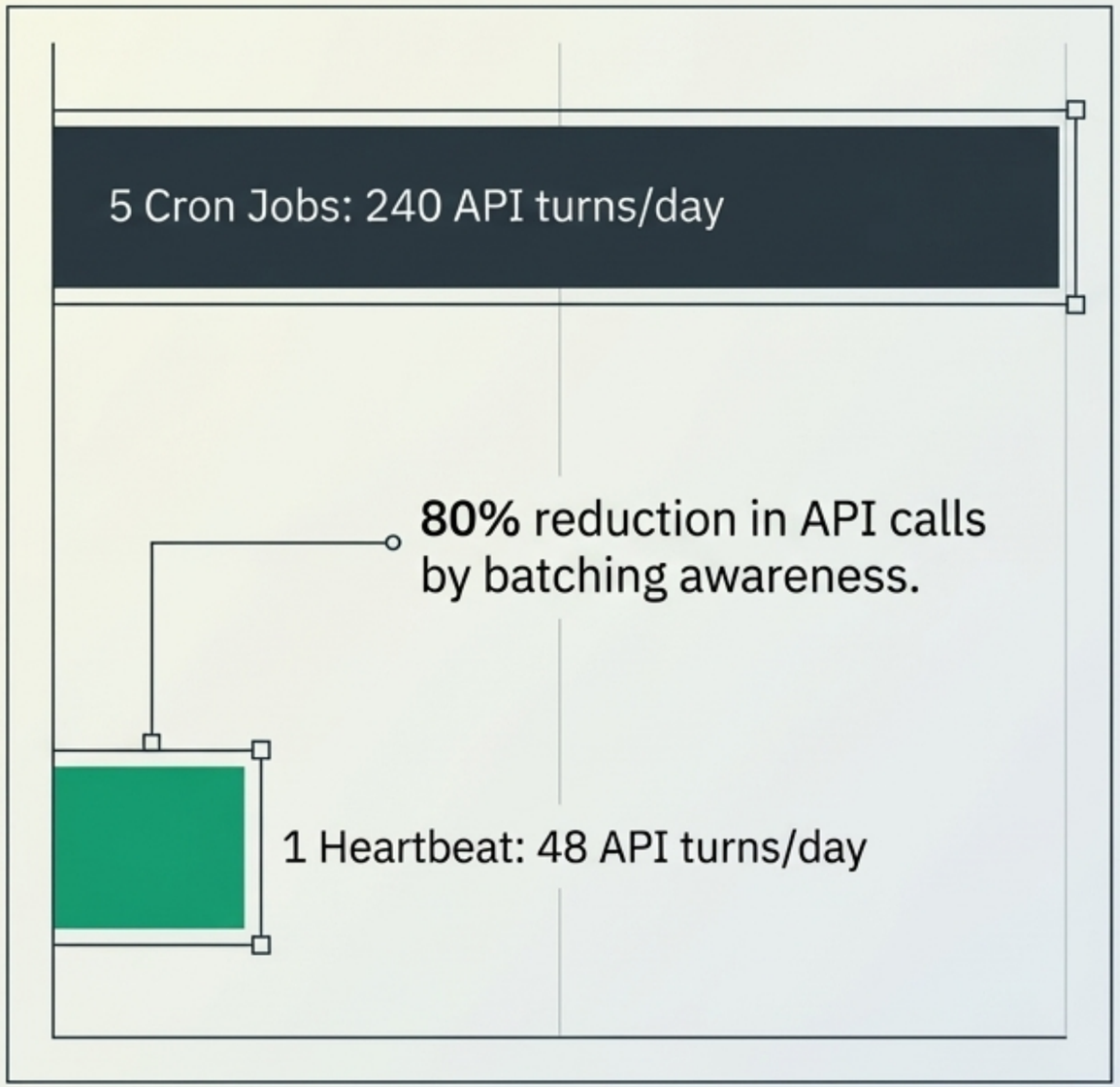
The Extension Hierarchy

A diagnostic matrix for capability design.

Skills (To KNOW)	MCP Servers (To ACCESS)	Plugins (To GAIN)
• **Format:** `SKILL.md` (Agent Skills Spec). Markdown + optional scripts. • **Role:** Teaches domain workflows or frameworks. Cross-platform standard. • **Code Required:** None.	• **Format:** External processes via stdio or SSE/HTTP. • **Role:** Connects external APIs, live data, and databases to the tool context. • **Code Required:** JSON config (`.mcp.json`).	• **Format:** Native (TypeScript inside gateway) or Bundle (file-based extensions). • **Role:** Alters core gateway capabilities (new providers, custom security hooks). • **Code Required:** TypeScript for Native; None for Bundle.

Escalation Path: Start simple with Skills, escalate to MCP, write Plugins only when modifying the Gateway.

Autonomy & Economics: Heartbeats vs. Crons



Cron Jobs (Precise Delivery)

- **Use Case:** Exact timing required (e.g., Send 9:00 AM summary).
- **Cost Math:** 5 jobs × 2 runs/hour × 24h = 240 isolated API turns/day.

Heartbeats (Periodic Awareness)

- **Use Case:** Batched checklist monitoring (HEARTBEAT.md).
- **Cost Math:** 1 heartbeat × 2 runs/hour × 24h = 48 batched API turns/day.

The Suppression Pattern

The gateway silently drops any heartbeat response containing the exact `HEARTBEAT_OK` token. The agent stays awake, watches the infrastructure, and costs zero notification noise.

Modality as UX Design: Voice TTS

1. off (Text Only)

Standard text generation. Zero voice processing.

2. always (Naive Processing)

Every single output goes through TTS. Naive execution causes lists, passwords, and code blocks to become uncopyable audio. Causes high user friction.

3. inbound (Production Default)

The gateway acts as a mirror. Text input yields text output; voice input yields voice output. The safest automated setting.

4. tagged (Agent as UX Designer)

The agent autonomously evaluates its response payload. It injects `[[tts]]` tags for narrative descriptions, while utilizing standard text for reference numbers and code.

The agent becomes the UI. Voice is used for description, text is used for data extraction.

Multi-Agent Orchestration

Shared Infrastructure (Gateway Level)

- Plugins (TTS, Voice, Search)
- Network Adapters & Channel Bindings (WhatsApp, Discord)
- Model Provider Configurations

The Routing Boundary

Pod A: Main Agent

Isolated State (Workspace Level)

- Identity (IDENTITY.md, SOUL.md)
- Memory (MEMORY.md, Daily Logs)
- Installed Skills

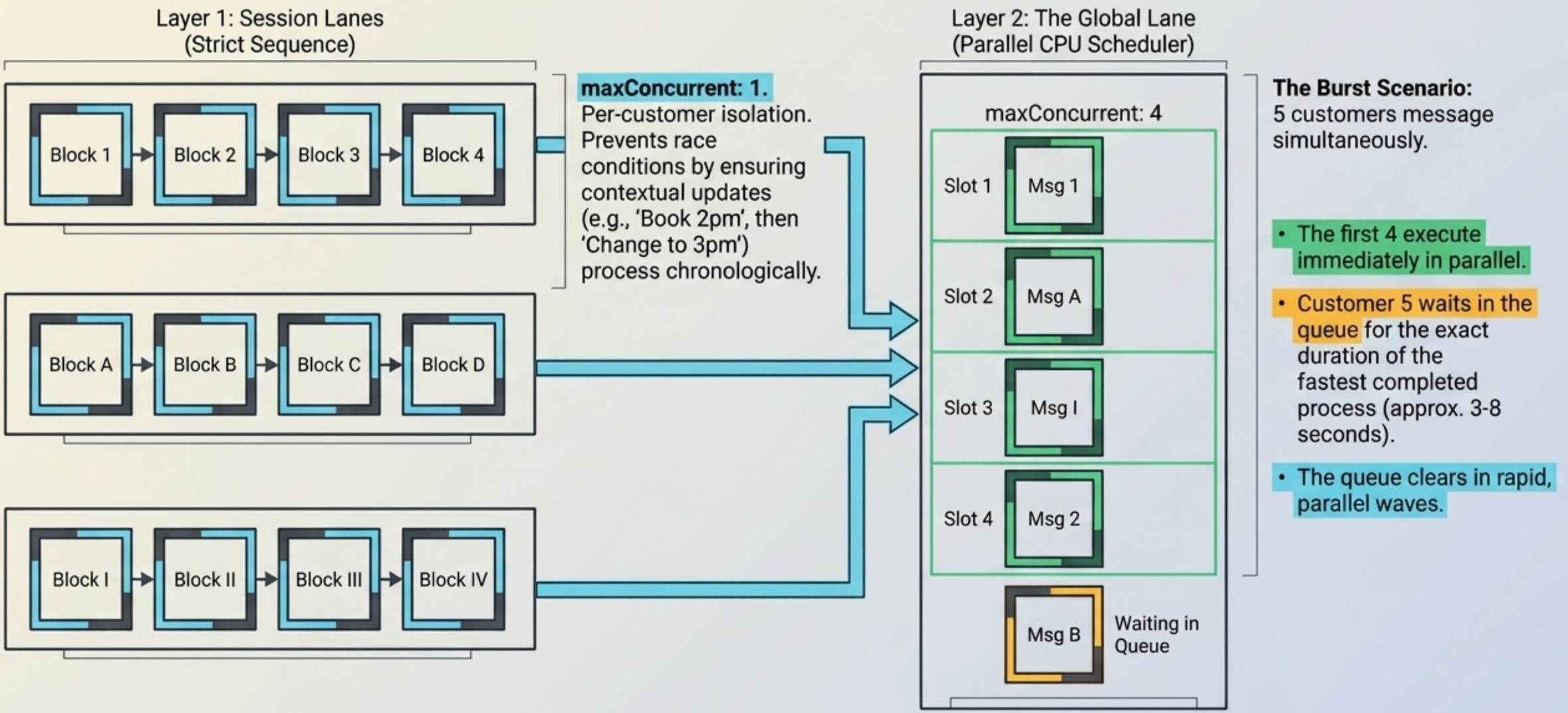
Pod B: Helper Agent

Isolated State (Workspace Level)

- Identity (IDENTITY.md, SOUL.md)
- Memory (MEMORY.md, Daily Logs)
- Installed Skills

Note: Peer-specific bindings bypass channel-wide rules. Identity updates require a session boundary (cache flush) to take effect.

The Concurrency Queue: Two-Layer Processing



The Tooling Reality: Google Workspace Integration

The Productivity Bridge (gog)

```
$ gog auth add --services calendar --readonly

Connecting OpenClaw to real data...
✅ Gmail: search, read, send
✅ Calendar: list, create
✅ Drive: search, upload
```

Connecting the Agent Loop to real, external data systems.

The Security Reality

Before gog:
- Agent manipulates self-generated text.

After gog:
- Agent manipulates your **actual inbox**, **password resets**, and **shared documents**.

Least Privilege Application: **Never** grant full access blindly. **Use strict scoping** (e.g., **--readonly**) to contain the blast radius of malicious skills or prompt injections.

Security Architecture (Tiers 1 & 2)

Tier 1: Tool Profiles
(The Binary Gate)

Binary Enforcement: Enforced in the Node.js runtime. Setting profile to '`coding`' allows shell execution; setting to '`messaging`' explicitly removes runtime/file capabilities.

Tier 2: Plugin Hooks
(The Human Doorbell)

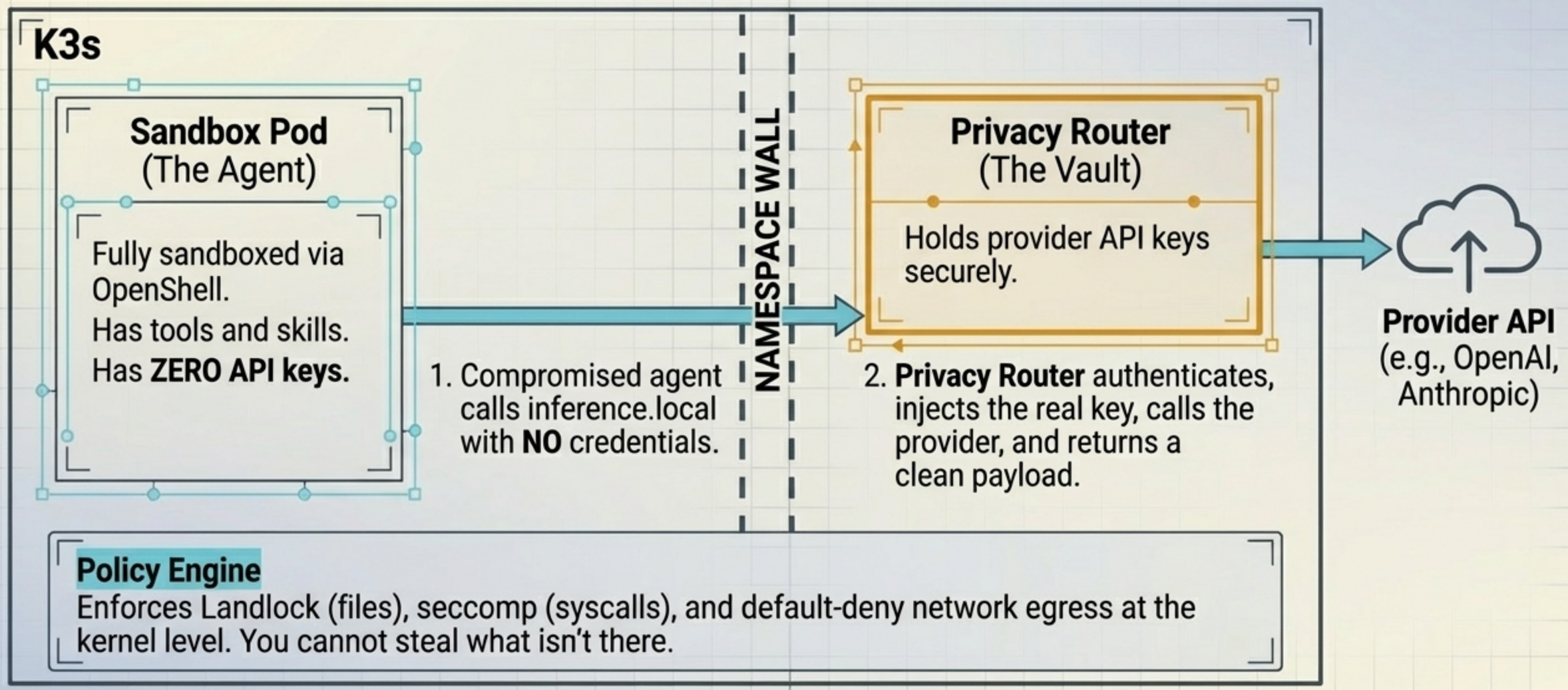
- **Intercepting Intent:** A custom TypeScript plugin utilizing `api.on('before_tool_call')`.
- The `requireApproval` **State:** Pauses tool execution, alerts the operator via WhatsApp/Slack, and awaits '`allow-once`', '`allow-always`', or '`deny`'.
- Fails closed on a 120-second timeout.

The Gap

MCP tools bypass this pipeline entirely.
MCP servers must self-protect.

Security Architecture (Tier 3): NemoClaw

Out-of-Process Physical Isolation.



Deployment Decision Matrix

Option A: Docker Compose (The Baseline)

- **Architecture:** Single VPS, single process. In-process key storage. Loopback bind + SSH tunnel.
- **Cost:** $5/mo infrastructure (+ Model costs).
- **Use Case:** Single operator, personal automation, low-risk environments.

Option B: NemoClaw (Zero-Trust)

- **Architecture:** K3s Cluster. 4-pod out-of-process policy enforcement.
- **Cost:** $15/mo infrastructure (+ Model costs).
- **Use Case:** Multi-tenant environments, high-risk operations.

Trust-Boundary Upgrade Triggers

- ☑ Adding paying customers.
- ☑ Compliance audit requirements.
- ☑ Bringing on a second operator who shouldn't have raw API key access.

Note: When any of these triggers are met, escalate from Docker Compose to NemoClaw.

Production Readiness: The Synthesis

Bridging the gap from configuration to production deployment.

7 Conditions for Calibrated Confidence:

1. ✅ **Dedicated Hardware:** Separate phone number for WhatsApp to prevent session/credential corruption.
2. ✅ **Capable Model:** `Claude Sonnet` or `GPT-4` tier to eliminate tool hallucination.
3. ✅ **Instructed Isolation:** Explicit prompt-layer limits on customer memory overlaps.
4. ✅ **Operator Gates:** `requireApproval` active on all customer-contacting tools.
5. ✅ **Zero Criticals:** `groupPolicy` set to `allowlist`; credentials secured (700 permissions).
6. ✅ **Continuous Uptime:** `Docker Compose` (minimum) or `NemoClaw` (for multi-tenant trust).
7. ✅ **Log Observability:** Active monitoring of the `Gateway log` to catch silent failures.

The Progression is Complete: The platform works when constraints are respected. You have evolved from User -> Extender -> Builder.